Demographic characteristics of locals and jetsetters: A Study of Extensibility using the Neighborhood Connectivity Survey

Xiaofan Liang, Seolha Lee, Hanzhou Chen, Benjamin de la Pena, Clio Andris

Abstract

Individuals connect to sets of places through travel, migration, social interactions, and telecommunications. This set of multiplex network connections comprises an individual's extensibility, a human geography term representing geographic reach. Here, we attempt to uncover the demographic and behavioral factors that correlate with high or low extensibility. We used a dataset of 950 individuals' self-reported surveys, and classify individuals into one of four different typologies: 1) hyperlocal, 2) majority local, 3) glocal, or 4) regional global patterns. We visualized individuals by their connection distances, strengths, types, and the diversity of connection destinations. We also tested whether these typologies correlate with local social support, ability to leverage social networks for disaster evacuation, frequency to travel or migrate between cities, and sociodemographic characteristics.

We found that respondents who are white, married, and have higher educational attainment were significantly associated with the glocal pattern, while those who reported as Black/African American, single, and having high school (or less) educational attainment tend to have more local social and spatial ties. We also found that glocal individuals are more likely to travel or resettle across cities, enjoy higher local social support, and have more evacuation options via social networks than individuals with mostly local ties. Our findings can help policymakers understand how individuals may be likely to exhibit different types of extensibilities, and how these factors can be used as 'rules of thumb' for estimating who may have distant or nearby connections.

Introduction

People operationalize their social life through connections to a set of places. These places can be people's childhood cities, locations of their families, regions of which they subscribed information, and locales where they have organization membership. Some individuals have many of these places, and some have few; some have distant connections and some have nearby connections. We might call someone a 'jetsetter' if they connect to a variety of places or perhaps a 'homebody' if they tend to enjoy having their ties and their energies invested in local places. These behaviors can be encompassed under the scholarly term *extensibility*, defined as the reciprocal of time-space convergence (Janelle, 1973; Adams, 1995), the geographic spread or reach of an agent (Adams, 2009), or geographic reach of a place or event (Kwan, 2000). When we map a place or an individual's extensibility, it can create an ego-centric network (Stutz, 1973) that links a place or an individual to all other places it connects to (e.g., where a person commutes in daily life). Such geographic reach can be considered as any connections with geographic space that allow us to leave the home, virtually, through others, or through movement.

Then how have people characterized extensibility in geographic space? Past research on extensibility, or more broadly, spatial social networks, tends to aggregate individual networks to places and associate the types of places with sociodemographic data. For instance, Facebook friendship data tells us that for a resident of Kentucky, the probability of having a Facebook friend outside 500 km is much lower than for a resident of Los Angeles (Bailey et al. 2018). Counties with higher average income, social capital, social mobility, and education also lead to a more extensive social network (Bailey et al. 2018). A study of British telephony call data also finds that wealthy locales have connections to many places, whereas poorer locales have few connections (Eagle, Macy & Claxton, 2012). Urban communities also tend to have more distant ties (Illenberger et al. 2011; Kowald et al. 2010), while rural communities tend to be local, centered by kinship (Fischer, 1982).

However, these insights are often place-based and rely on homogenous ties from a single data source. As such they do not provide the full story of individual extensibility and its interactions with other demographic and behavioral factors. Neighborhood populations are becoming more transient and heterogeneous so that their social networks within the same geographic boundaries can vary (Mazumdar et al., 2018; Webster et al., 2017). While social network researchers have found similar conclusions on the individual level (e.g., rich and educated individuals also have more dispersed social networks) (Eijk, 2010), they rely on prior assumptions from sociological theories and simple distance metrics to characterize the networks and often focus solely on social relationships. Different types of connections can play complementary or contradicting roles in an individual's life across various dimensions, such as mobility (Larsen et al, 2006; Kowald, Axhausen, 2012; Picornell et al., 2015), health (Perkins et al., 2015; Chan et al., 2011), political orientation (Boutyline, Willer, 2016; Wang et al., 2020), and communication (Calabrese et al., 2011) etc. They may also distribute differently at various spatial scales, though we only know the difference for social relationships (e.g., friendship vs. kin) from Boessen’s study (2014).

Besides sociodemographic attributes, the structure of personal social networks is also correlated with individuals’ social support, residential mobility, and travel behaviors. For example, Viry (2012) found that people’s social support (i.e., the number of supporting ties) are not affected by the geographic distribution of the networks and the frequency of moving, though those who move frequently lean toward a sparsely knit and transitive social network. Transportation planners have also used social networks as a mechanism to explain and predict leisure travel destinations (Kowald & Axhausen, 2015), travel demands, and mode choices (Pike & Lubell, 2016). Thus, we can expect various extensibility patterns may have distinguishable effects on people’s local social support, intercity travel frequency, and resettlement frequency.

Characterizing (i.e. measuring) ego-centric spatial social networks with multi-modal connections can be challenging. First, we lack a dataset that possesses both in-depth information about the individuals (e.g., socio-demographic information) and the types of connections, which hinders our understanding of their interactions. Also, individual extensibility patterns are difficult to characterize due to the multiplicity of nodes and edges. An individual’s ego-centric network can include attributes that are aggregated statistics (e.g., the number of connections), categories (e.g., the types of connections), and distributions (e.g., distribution of connection distances).

Therefore, the challenge becomes how to maintain the richness of all these dimensions while reducing them in a comparable format. Lastly, very little can be leveraged from traditional metrics in social network analysis(e.g., centrality measures) for the ego-centric networks in this study since they lack interconnectivity. The common measures of social network analysis, such as centrality, modularity, etc., are defined by the relations of a node with the entire network, while the egos collected by surveys are usually not connected to each other by links. Limited studies have done classification of the egos of such ego-centric networks with connectivity characteristics (Andris, 2016). Thus, we need to create other metrics for disconnected ego-centric networks.

We contributed to existing literature by creating a new dataset of ego-centric and multi-modal spatial social networks and characterizing such networks through a data-driven machine learning model. We asked what are the common configurations of individual extensibility and whether each pattern is statistically related to individuals' demographic and behavioral attributes. To answer these questions, we clustered 950 individuals (with more than 20,000 connections) into four groups that are distinctive in distance distribution and link richness. Then we used post-hoc tests of ANOVA and Chi-square to reveal whether these groups can be distinguished by a-prior sociodemographic and behavioral factors. Our results suggest correlations between connectivity patterns and race, education, relationship status, intercity travel frequency, local social support, and resettlement frequency, but not in political orientation, age, gender, children status, and employment. Since individuals' connectivity data are hard to collect consistently, these correlations can help determine policies that are contingent on connections and social capitals, such as which groups are more likely to travel between cities for public health control, who are least likely to evacuate due to the lack of ties outside of communities for disaster relief, and who may need local social supports for community health.

Data and Methods

Data: Neighborhood Connectivity Survey

Our study uses data collected from the Neighborhood Connectivity Survey (NCS), a large mail-based survey conducted in 2017 and 2018. A mailing was sent to participants selected from cities near three major locales: the Akron, Ohio Metropolitan Area; the State College, Pennsylvania Metropolitan Area; and Philadelphia County, Pennsylvania, i.e. urban Philadelphia. In 2017 and 2018, We mailed 20,000 mailings, and received 1023 surveys, while 940 are sufficiently completed for our research purposes. The survey includes four modules: connectivity, social life, behaviors, and demographic metrics, which took roughly 30 minutes to finish. Participants could answer the survey on paper or online and were rewarded with a gift card to nationwide retailers for their participation.

Variables: Connections, Demography, and Behaviors

We define connectivity as individuals' connections to geographic locations. To protect privacy, locations are reported on the level of cities (and some international links reported countries). We asked thirteen relational questions and grouped them into five categories: migration (i.e. where people have lived for an extensive period of time), social ties (e.g. close friends/families, communication, financial/legal supports, etc.), institutions (e.g., school, affiliated organizations), news (i.e., subscriptions to non-local news), and travels (i.e., where people have visited). These connectivity answers could be presented as an egocentric network centered at a respondent's home location and connected to geographic locations to which the individual has connections. 950 responses out of 1023 total responses had reported more than two connections and 10 out of 950 responses missed sociodemographic information. Thus, in this study, we used 950 responses for connectivity classification and 940 for statistical analyses.

Demographic variables include age, race, employment status, gender, relationship status, political orientation, and education level. For the 940 respondents, 617 of them are females and 288 are males (35 reported other or did not disclose gender). 79.2% (n=752) of respondents were White/Caucasian, 12.0% (n=114) were Black/African American, 1.89% (n=18) were Hispanic/Latino and another 1.89% (n=18) were Asian, 2.32% were bi-racial (n=22). Two respondents were Middle Eastern/North African, three were Native American, and 1.16% (n=11) did not disclose this information. Most respondents were employed (n=549), about half were married (n=454) and about half did not have children in the home (n=473) (See S.I. for survey questions and demographics).

Behavioral factors include local social support index, intercity travel frequency, resettlement frequency, and the percentages of people evacuated to locations of close friends and families during disasters. Local social support index is generated based on questions about people's social life, such as how often they have lunch with coworkers and how many friends they feel comfortable to have dinners with. The index scales from zero to one, representing low to high levels of local social support. We derive an estimate of people's intercity travel frequency, based on how many times they used intercity modes of transport (e.g., flights, intercity buses etc.). Resettlement frequency counts the number of cities that people have lived in for more than six months in the past. We also asked people to fill in locations they will go for shelter if a disaster happened in the local areas for two weeks, two months, and forever. We then compare those locations to locations of their close friends and families to calculate the percentages.

Clustering using Machine Learning - K-means

We chose unsupervised learning to overcome the limitations of a-prior assumptions of connectivity patterns. Machine learning techniques have been widely used to study network-based data with different purposes: finding a prevalent subgraph pattern (Diane et al., 2007), classifying or identifying different members (nodes) from a communication/social networks (Nurek, Michalski, 2020; Alsayat and El-Sayed, 2016), or measuring dynamics in networks (Agarwal, Bharadwaj, 2015).

There are several advantages of using unsupervised learning in social network study: first, the algorithms allow us to use multidimensional data for classification, which is a common case in heterogeneous networks. Also, we don't have the number of clusters apriori of classification experiments and the learning algorithms suggest an optimal number of clusters as well.

Prominent clustering algorithms based on machine learning are nearest neighborhoods algorithms (e.g. K-means), decision tree algorithms (e.g. hierarchical clustering), and model-based clustering. We tested and compared the results from three algorithms using most exemplary r-packages for each algorithm. The input data used for all three algorithms were the same, which will be illustrated in the next paragraph. We chose K-means clustering for the final clustering experiment since the algorithms resulted in an adequate number of clusters for further analyses and had better internal consistency in each cluster compared to other algorithms. See Appendix for further details of the clustering experiment and results.

K-means algorithm has feature vectors as its input and clusters the samples based on the distance between the vectors (euclidean distance in most cases) in the vector space. The algorithm iterates assigning clusters to samples until the sum of the distances between the samples in each cluster reaches the minimum. We have eight input variables for the algorithm to characterize each individual's egocentric networks. Five concerns with distance distribution of nodes, while the other three are the total number of links, the number of unique places connected to the ego, and the number of connection types. They represent network structure's spatial scales, volumes, and diversity respectively.

To convert the distance distribution into a vector, we cut the distribution into 5 distance bins: <5km, 5-50km, 50-1300km, >1300km, and non-US. The thresholds were selected based on the observed distribution, such as visually distinctive troughs (5, 50km) or natural breaks (1300km), and can be interpreted as connections in the neighborhood, city, and regional scale (see Bossen et al., 2017 for a similar application). Especially in our work, distance also can imply connections to certain regions because the home locations of our respondents are all located in the midwest/northeast boundary (see Figure 1).

To prevent the total number of links from overly driving the clustering result, we used the percentage of links that fall in each distance bins instead of the absolute numbers. Also, to avoid any feature dominating the clustering, we scaled the three other features between 0 to 1, by dividing each value by the maximum values (13 for the link types, 38 for unique places, 64 for the total number of links).

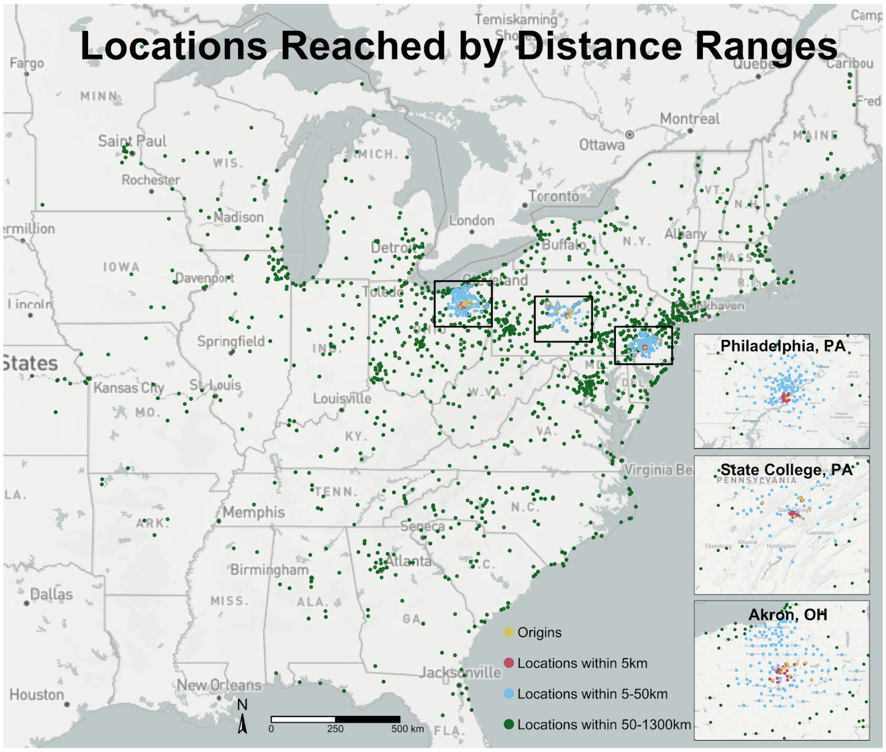

Figure 1: Destinations in different distances range from our origin cities.

Statistical tests with Chi-Square and ANOVA.

To examine whether the extensibility clusters have statistically distinctive demographic and behavioral characteristics, we used Chi-Square post-hoc tests for all categorical variables in demography and ANOVA post-hoc tests (Tukey HSD) for continuous variables in behavioral factors. We calculated the standardized residuals in Chi-Square post-hoc tests for each cluster. The residuals represent the extent of which the observed counts of a demographic category in a cluster deviates from the expected counts (i.e., total counts divided by the number of clusters) normalized by cell variance (Agresti, 2007):

$$Std\ Residuals\ =\ (observed\ -\ expected)/sqrt(V),\ where\ V\ is\ the\ residual\ cell\ variance$$

We also used bonferroni correction for the p-values to account for the multiple comparisons. ANOVA post-hoc tests conduct pairwise comparisons between the clusters on the variables. We chose Tukey HSD to report the statistical significance of the mean difference, as it is recommended for groups with unequal sample size, which is the case in our survey.

Results

Classification

The K-means clustering returns four clusters. Each cluster has a distinct feature distribution (see detailed statistics in S.I. Table 1). We called the first cluster *Hyperlocal*, because the majority of the connections are concentrated within five kilometers from their home locations

(Figure 2). These connections tend to be social and institutional ties and have very few non-local news subscriptions and travel outside of the local areas, indicating a close-knit local social circle (Figure 3). The 195 people in this category have very likely been living in the same city until now as the distribution of spatial ties highly overlap with local social ties. Consistent to this interpretation, the number of unique places they are connected to is also the lowest compared with people from other clusters (Figure 2).

The second cluster is called *Metropolitan*, named after the concentration of links which falls within the size of a metropolitan area (i.e. 50km) (Figure 2). 235 people fall in this category. The distance distribution of people's migration history closely follows their social and institutional ties (Figure 3) at both the neighborhood (0-5km) and the city (5-50km) range. People in this cluster also enjoy a decent number of total connections and connection types as those in *Hyperlocal*.

The third cluster has the highest average number of total connections and mixed-distance ties, thus called *mixed-many* (Figure 2). The 292 people in this cluster establish local connections through institutions, while at the same time, maintain extensive social networks and spatial footprints (migration and travel) (Figure 3). Besides topping the total connections, this category also has the highest percentage of connections to international destinations and the most diverse ties in terms of connection types and the number of unique places. We expect that some people in *mixed-many* are privileged and resourceful given the glocal pattern of connectivity.

Last but not least we have *regional-few*, a cluster that is featured by the fewest number of total links but most of which extend across regions (Figure 2). The 228 people in this cluster lack local ties and have the least diverse connection types. While their institutional connections are mostly local, people's spatial, social, information (news), and travel histories and networks overlap at a regional distance range that is greater than 50km but smaller than 1300km (Figure 3). The overlap may come from the fact that people in this cluster have recently moved to the current city (e.g., college students) and still maintain a social life from their original places. We can also interpret these people as lonely wanderers that have been to a few cities but are not deeply rooted.

Figure 2: Box-plots and sample stars. The boxplot shows the descriptive statistics of each feature in each cluster. All values are scaled between 0 and 1. The sampled egocentric networks on the right show how the individuals are actually reaching to the locations in the geographic space. The edges are weighted by the number of links connected to each location. Though the respective shapes vary, the geographic extents where the majority of their links fall in are similar in each cluster.

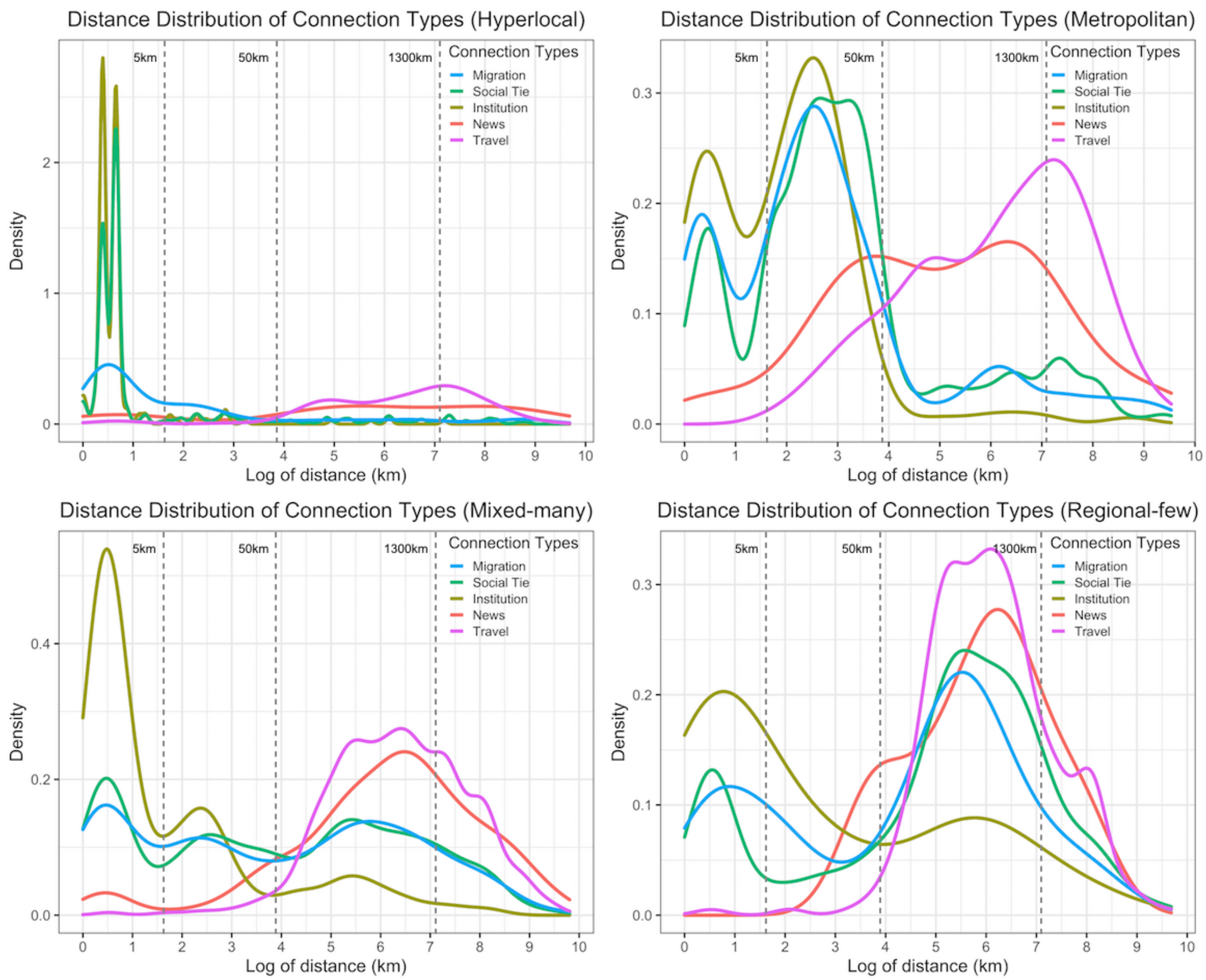

Figure 3: Distance distribution of various connection types for each cluster.

Statistical Correlation with Sociodemographic and Behavioral Characteristics

Table 2 reported the standardized residuals from chi-square post-hoc tests. Age, employment, gender, children status, and political orientation variables are relatively well-distributed across the clusters and thus do not exhibit significant correlation with one or more patterns.

We found that people who have high school education or less are statistically more likely to have locally concentrated ties, as featured by the *Hyperlocal* and *Metropolitan* patterns. Conversely, people who have a Bachelor's degree or above are more likely to have a *mixed_many* network pattern. We also observed that pursuing an associate degree can help motivate people to expand their spatial social networks beyond the local context. We postulated that education beyond high school may have a significant impact for people to meet others that came from places far away or to go to places outside of their comfort zones.

People who identified as white also disproportionately concentrated in the *Metropolitan* and *Mixed_many* clusters, while Black or African American tend to have *Hyperlocal* style of personal networks (with a residual of 6.86). When converted to real numbers, 45% (n=51) of the Black or African American respondents have their extensibility patterns classified as *Hyperlocal*, which is much higher than the expected 25% (one in four patterns). In addition, race and education levels are correlated and interactive to keep the black communities in local places. Decomposing education by race in *Hyperlocal*, we found that only 57% (n=29) of the Black or African American in the *Hyperlocal* were educated in high school or less, which means the rest of those who received more education may still maintain a locally-oriented extensibility. Decomposing the race of those who received associate degrees across four clusters, we found that Black or African American are more likely than White American (45% vs. 17%) to be in *Hyperlocal*. Such disparity indicates that the effect of education at expanding one's spatial social networks is less on Black or African American, which may be due to the personal preferences or financial constraints to go to schools in their home cities. The difference may also be a population characteristic to have close-knit relationship circles on the neighborhood level.

Besides race and education, people who identify as single seem to concentrate in the *Hyperlocal* cluster too, but this effect may be explained by education levels. All single people in *Hyperlocal* have an education level of high school or less. In contrast, people who are married tend to have a *Mixed_many* type of connectivity pattern. The marriage status for the non-white respondents correspond well with education levels (the nine non-white respondents (5.5%) in the *Mixed_many* have all received associate degrees or above), while less so for the white respondents (only 80% have received associate degrees or above). Thus, we expect that Black or African American who are single and less educated are particularly subjective to the *Hyperlocal* connectivity pattern. Higher education is the key for the non-white population to expand their extensibilities.

Table 1: Standardized Residuals from Chi-square Post-hoc tests.

Sociodemographic Vars	**Hyperlocal**	**Metropolitan**	**Mixed_many**	**Regional_few**	**Count**
Age: 18-24	1.03	-0.7	-0.99	0.8	41
Age: 25-34	1.34	-0.97	-0.9	0.68	165
Age: 35-54	0.44	0.45	-1.42	0.67	137
Age: 54-65	-1.15	0.89	-0.04	0.24	141
Age: 65+	-1.07	0.08	2.46	-1.73	www
Employment: Unemployed	2.04	0.85	-2.48	-0.08	47
Employment: Retired or Disabled	1.03	1.45	-0.17	-2.25	282
Employment: Student	0.19	-1.58	0.3	1.1	42
Employment: Employed	-1.97	-1.08	1.15	1.69	541

Gender: Female	0.03	-0.07	2.56	-2.72	617
Gender: Male	-0.03	0.07	-2.56	2.72	288
Education: High school or less	**6.29****	**4.46****	**-6.49****	**-3.31***	394
Education: Associate	**-3.8****	0.1	2.46	0.77	259
Education: Bachelor	-1.64	**-3.64****	**3.65****	1.24	169
Education: Master or above	-2.67	**-2.93***	2.39	2.84	81
Political Orient: Very right	-1.2	-0.04	0.18	0.92	58
Political Orient: Moderate right	-1.07	0.63	0.1	0.21	148
Political Orient: Neutral	1.35	1.6	-1.69	-0.95	231
Political Orient: Moderate left	-0.46	-0.8	1.65	-0.61	219
Political Orient: Very left	0.85	-1.61	-0.16	1.02	137
Race: White or Caucasian	**-7.21*****	**3.26***	**4.34*****	-1.2	752
Race: Black or African American	**6.86*****	**-3.3***	**-3.77****	0.94	114
Race: Other	2.3	-0.79	-1.86	0.64	63
Relationship: Single	**4.26*****	0.15	**-4.53*****	0.73	186
Relationship: In a relationship	0.06	1.57	-0.51	-1.09	121
Relationship: Married	**-3.71****	-1.63	**3.33***	1.55	454
Relationship: Divorced or separated	-0.53	0.28	0.35	-0.16	108
Relationship: Widowed	1.18	0.55	0.8	-2.55	65
Children below 18: Yes	0.03	0.01	-0.06	0.02	212
Children below 18: No	-0.09	0.02	0.07	-0.01	473

Note: *p<0.05; **p<0.01. ***p<0.001. P-value is adjusted by Bonferroni correction. The standardized residuals should be interpreted across sociodemographic subtypes (e.g., male and female) for a particular cluster. A statistically significant standardized residual means that a sociodemographic attribute is highly concentrated in a cluster beyond expected mean (see Method for equation).

In terms of the behavioral characteristics, the ANOVA post-hoc tests report statistically significant mean differences between two clusters. People who have more long-distance connections (e.g., the *Mixed_many* and *Regional_few*) travel more often between cities. The correlation is reasonable because connections provide motivations (and evidence) for people's travels in the past, such as visiting families or going to alumni events.

People with *Hyperlocal* and *Metropolitan* style of extensibility also reported less local social support than people in the *Mixed_many*, despite the former having a high concentration of local ties. Since local social support index only measures the quality of social life locally, the result indicates that people in *Mixed_many* are more likely to receive social support from their local networks than people in *Hyperlocal* and *Metropolitan*, even if they share a similar number of total connections (the mean is 14, 10, 12 in *Mixed_many, Metropolitan,* and *Hyperlocal* respectively).

The four clusters also have a statistically distinguishable resettlement frequency. From high to low, people in *Mixed_many* have lived in most places in the past, then to people in *Metropolitan, Regional_few,* and lastly *Hyperlocal*. If we look at the moving frequency with the classification results (see Figure 2 and 3), the moving frequency helps explain why people in *Mixed_many* have spatially distributed personal networks across long distances. People in the *Metropolitan* cluster may have moved a few times in the same metropolitan areas, while people in the *Regional_few* may have just resettled in the current city far from where they used to live and have not invested in the local networks. People in *Hyperlocal* are most likely to stay in the same cities and thus focus on cultivating local and neighborhood ties.

Lastly, we tested whether people with different extensibility patterns will react differently to evacuation scenarios. We did not find the distances to preferable evacuation locations differ significantly across the clusters, but found that *Hyperlocal* has the fewest percentages of people (36%) that will evacuate to locations of closest friends and families. We do not investigate why people in the *Hyperlocal* cluster chose other locations for evacuation, but we believe that they did not reach out to closest friends and families because they are likely to be impacted by the same disasters due to co-location. As such, people with *Hyperlocal* pattern may lack critical social support and resources to relocate for disaster relief outside of their home locations. In contrast, 84% of *Mixed_many* respondents identified plausible evacuation locations, while only 45%, 43%, and 27% of people in *Hyperlocal, Metropolitan,* and *Regional_few* have filled in any cities to go for evacuations. Thus, the extensive spatial social networks in *Mixed_many* are advantageous at providing more options and support during disasters, while the lack of close social ties outside of communities in *Hyperlocal* may hinder temporary relocation.

Table 2: ANOVA Multiple Comparisons: Tukey HSD

Dependent Variable	Cluster (a)	Cluster (b)	Mean Difference (a-b) and Confidence Intervals
Intercity Travel Frequency	Mixed_many	Hyperlocal	17.53 (4.28, 30.79) **
Intercity Travel Frequency	Mixed_many	Metropolitan	25.08 (12.47, 37.69) ****
Intercity Travel Frequency	Regional_few	Metropolitan	16.92 (3.52, 30.32) **
Local Social Support	Mixed_many	Hyperlocal	0.04 (0.01, 0.08) *
Local Social Support	Mixed_many	Metropolitan	0.02 (0.02, 0.08) ***
Resettle Frequency	Metropolitan	Hyperlocal	0.74 (0.14, 1.33) **
Resettle Frequency	Metropolitan	Regional-few	0.79 (0.14, 1.44) **
Resettle Frequency	Mixed_many	Hyperlocal	2.45 (1.93, 2.98) ****
Resettle Frequency	Mixed_many	Metropolitan	1.72 (1.21, 2.23) ****

Resettle Frequency	Mixed_many	Regional-few	2.51 (1.94, 3.09) ****
Percentage that evacuates to locations of closest friends and families	Metropolitan	Hyperlocal	0.37 (0.18, 0.55) ****
Percentage that evacuates to locations of closest friends and families	Mixed_many	Hyperlocal	0.36 (0.21, 0.52) ****
Percentage that evacuates to locations of closest friends and families	Regional_few	Hyperlocal	0.29 (0.06, 0.53) **

Note: *p<0.05; **p<0.01; ***p<0.001. ****p<0.0001. P-value is adjusted by Bonferroni correction. Only significant results are shown. The Values in the parentheses are confidence intervals.

Conclusion

This study created ego-centric networks of individual connectivity through an extensive mail-based survey named Neighborhood Connectivity Survey. The survey provided a unique dataset that included a wide range of spatial social connections of individuals with their socio demographic information. Then, we conducted unsupervised clustering of the individual spatial social networks using the k-means algorithm to characterize the individual connectivity with multiple features. Lastly, we examined the tendencies in sociodemographic characteristics, social life, and spatial activities of individuals with each connectivity pattern through ANOVA and chi-square tests.

The study's major findings are four-fold: first, different types of links (migration, social ties, institutions, news, travel) had distinctive spatial distributions from our dataset. Second, our clustering method resulted in four distinctive types of individual spatial social networks: *Hyperlocal, Metropolitan, Mixed-many,* and *Regional-few*. The clusters vary largely in terms of the distance distribution of the links and are distinguished by the richness (total number, unique locations connected, and the number of different types of links) of the links.
Third, individuals in the four clusters showed different tendencies in their demographic characteristics, the level of local social support they received, the frequency of intercity travel and resettlement, and options for disaster evacuation. Among demographic variables, race, education, relationship status are correlated with individuals' spatial social network patterns, while age, gender, children status, employment, and political orientation didn't show a significant correlation with the clusters. It is notable that higher education attainment had the effect of expanding the spatial social network, yet the effect was less on Black or African American populations. Lastly, individuals with the connectivity pattern of *Mixed-many* had more intercity travel and higher local social support than typologies with most local ties (i.e., *Hyperlocal, Metropolitan*). Individuals in the *Mixed-many* group also have lived in most places in the past, which explains their connectivity patterns that are sparsely distributed across long distances.

The study has a couple of limitations that could be addressed in the future. First, the sample population was limited to residents in a few neighboring cities. Since the cities were concentrated near the Northeast region of the US, the distance distribution could be reasonably consistent across the sample population. It might not be the case if the sample includes more cities from different regions in the future. Also, there were still variances between individuals in each cluster. We focused on the mean values of the features used for clustering to characterize each connectivity pattern. Yet, the box plots still showed internal variances. Lastly, we didn't have a detailed explanatory mechanism for the clusters. Unsupervised clustering captures intrinsic tendencies but doesn't explain why. Some findings, like resettlement behavior, could provide partial explanations of how the individuals ended up having such connectivity patterns, but not enough to explain the whole.

We expect this study will lead to some future studies. First of all, examining direct correlation with a smaller number of variables from our survey data will provide a more in-depth understanding of how different connections are associated with demographic or lifestyle factors. Also, it would be desirable to compare the individual-based connectivity characterized in our study to place-based knowledge. The comparison will let us know the extent of diversity of individual connectivity within the same geographic boundary so that it can add meaningful insights to the existing discourse of place-based connectivity.

References

Adams, P. C. (1995). A reconsideration of personal boundaries in space-time. *Annals of the Association of American Geographers*, *85*(2), 267-285.

Adams, P. C. (2009). *Geographies of media and communication*. John Wiley & Sons.

Agarwal, V., & Bharadwaj, K. K. (2015). Predicting the dynamics of social circles in ego networks using pattern analysis and GA K-means clustering. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, *5*(3), 113–141.

Agresti, A. (2018). *An introduction to categorical data analysis*. John Wiley & Sons.

Alsayat, A., & El-Sayed, H. (2016). Social media analysis using optimized K-Means clustering. *2016 IEEE/ACIS 14th International Conference on Software Engineering Research, Management and Applications, SERA 2016*, 61–66.

Boessen, A., Hipp, J. R., Smith, E. J., Butts, C. T., Nagle, N. N., & Almquist, Z. (2014). Networks, space, and residents' perception of cohesion. *American journal of community psychology*, *53*(3-4), 447-461.

Boutyline, A., & Willer, R. (2017). The Social Structure of Political Echo Chambers: Variation in Ideological Homophily in Online Networks. *Political Psychology*, *38*, *551-569*.

Calabrese, F., Smoreda, Z., Blondel, V. D., & Ratti, C. (2011). Interplay between Telecommunications and Face-to-Face Interactions: A Study Using Mobile Phone Data. *PLoS ONE*, *6*(7), e20814.

Chan, A., Malhotra, C., Malhotra, R., & Østbye, T. (2011). Living arrangements, social networks and depressive symptoms among older men and women in Singapore. *International Journal of Geriatric Psychiatry*, *26*(6), 630–639.

Gomez-Rodriguez, M., Balduzzi, D., De, B. M., Schölkopf, B., & De, B. M. (2011). *Uncovering the Temporal Dynamics of Diffusion Networks*.

Janelle, D. G. (1973). Measuring human extensibility in a shrinking world. *Journal of Geography*, *72*(5), 8-15.

Kowald, M., & Axhausen, K. W. (2012). Focusing on Connected Personal Leisure Networks: Selected Results from a Snowball Sample. *Environment and Planning A: Economy and Space*, *44*(5), 1085–1100.

Kowald, M., van den Berg, P., Frei, A., Carrasco, J. A., Arentze, T., Axhausen, K., Mok, D., Timmermans, H., & Wellman, B. (2013). Distance patterns of personal networks in four countries: A comparative study. *Journal of Transport Geography*, *31*, 236–248.

Kowald, M., Arentze, T. A., & Axhausen, K. W. (2015). Individuals' spatial social network choice: model-based analysis of leisure-contact selection. *Environment and Planning B: Planning and Design*, *42*(5), 857-869.

Kwan, M. P. (2000). Human extensibility and individual hybrid-accessibility in space-time: a multi-scale representation using GIS. In *Information, place, and cyberspace* (pp. 241-256). Springer, Berlin, Heidelberg.

Larsen, J., Urry, J., & Auxhausen, K. W. (2006). *Mobilities, Networks, Geographies*. Ashgate Aldershot, UK.

Mazumdar, S., Learnihan, V., Cochrane, T., & Davey, R. (2018). The Built Environment and Social Capital: A Systematic Review. *Environment and Behavior*, *50*(2).

Cook, D., & Holder, L. (2006). *Mining Graph Data*. John Wiley & Sons, Inc.

Nurek, M., & Michalski, R. (2020). Combining Machine Learning and Social Network Analysis to Reveal the Organizational Structures. *Applied Sciences*, *10*(5), 1699. https://doi.org/10.3390/app10051699

Pike, S., & Lubell, M. (2016). Geography and social networks in transportation mode choice. *Journal of transport geography*, *57*, 184-193.

Perkins, J. M., Subramanian, S. V., & Christakis, N. A. (2015). Social networks and health: A systematic review of sociocentric network studies in low- and middle-income countries. *Social Science and Medicine*, *125*, 60–78.

Picornell, M., Ruiz, T., Lenormand, M., Ramasco, J. J., Dubernet, T., & Frías-Martínez, E. (2015). Exploring the potential of phone call data to characterize the relationship between social network and travel behavior. *Transportation*, *42*(4), 647–668.

Stutz, F. P. (1973). Distance and network effects on urban social travel fields. *Economic Geography*, *49*(2), 134-144.

Van Eijk, G. (2010). *Unequal networks: Spatial segregation, relationships and inequality in the city* (Vol. 32). Gwen van Eijk.

Viry, G. (2012). Residential mobility and the spatial dispersion of personal networks: Effects on social support. *Social networks*, *34*(1), 59-72.

Wang, C., Lizardo, O., & Hachen, D. S. (2020). Neither influence nor selection: Examining co-evolution of political orientation and social networks in the NetSense and NetHealth studies. *PLOS ONE*, *15*(5), e0233458.

Webster, N. J., Ajrouch, K. J., Tighe, L., & Antonucci, T. C. (2017a). RACE, GENDER, AND SOCIAL NETWORKS: THE ROLE OF NEIGHBORHOOD CONTEXT AND IMPLICATIONS FOR HEALTH. *Innovation in Aging*, *1*(suppl_1), 941–941.